DOCUMENT RESUME

ED 150 721 EA 010 353

AUTHOR Wright, E. N.; And Others
TITLE Issues Concerning the Design and Implementation of an Experimental Study of the Effects of Class Size.
PUB DATE Mar 78
NOTE 25p.; Paper presented at the Annual Meeting of the American Educational Research Association (Toronto, Ontario, March 27-31, 1978)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Class Size; *Educational Research; Elementary Education; *Research Design; *Research Problems; Socioeconomic Status; Teacher Background; Teaching Experience; Teaching Skills; Teaching Techniques
IDENTIFIERS Ontario

ABSTRACT

Previous research on class size has serious limitations because factors such as the abilities of pupils or the experience and teaching style of teachers were not controlled, and because of inconsistent criteria for large and small classes and for criteria of success. A major two-year study was mounted with four class sizes, random assignment of teachers and pupils to class sizes, and a variety of measures--observation, achievement, and attitudinal. The study was successfully implemented and completed because of the process of consultation and involvement followed. The paper details the design and the process. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED150721

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ISSUES CONCERNING THE DESIGN AND IMPLEMENTATION OF AN EXPERIMENTAL STUDY OF THE EFFECTS OF CLASS SIZE

BY

E. N. Wright
Gary Eason
John FitzGerald
Stanley M. Shapson

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

E. N. Wright

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

A paper prepared for the 1978 Annual Meeting of the American Educational Research Association, Toronto

Session No. 3114

EA 010 353

ISSUES CONCERNING THE DESIGN AND IMPLEMENTATION OF AN EXPERIMENTAL STUDY OF THE EFFECTS OF CLASS SIZE*

E. N. Wright
Gary Eason
John FitzGerald
Stanley M. Shapson**
The Board of Education for the City of Toronto

The "class size" question has been resurrected again as a result of falling enrolments, changing social values, changing public perceptions of the role of the teacher, and debates about the "quality of education." Almost every ten years as enrolments change, class size seems to be the subject of another "major" study and of many minor ones. As far back as 1932 Carr wrote that no new study of this subject should be undertaken without a clear statement of the objectives of education as a background to which the effects of changing the class size variable could be related. In spite of such warnings, and partly because of the morass of conflicting research findings in the literature, a new study of the question was initiated at the request of the Toronto Teachers' Federation and funded by the Ministry of Education for the Province of Ontario.

This paper has three parts: the first discusses the reasons the present study was done, and issues which were considered in the research design; the second describes the conduct and design of the study; the third discusses how the problems described in the first part of the paper were managed. Following this paper are two papers presenting other information from the same study (Shapson, Wright, Eason and FitzGerald, 1978; FitzGerald, Wright, Eason and Shapson, 1978).

---

* This research project was funded under contract by the Ministry of Education, Ontario.

** Now at Simon Fraser University.

PART I

In order to understand more fully the implications of research questions related to the class size study, it would be helpful to review briefly the social climate in the Province of Ontario in 1970-71. It was known that the population projections indicated that for the next several years the province would experience a decline in the number of pupils entering the school system; that one consequence of this would be that there would be a threat to teachers' job security, proportionate to the decline in enrolment. Also, the provincial government had embarked on a policy of restraining the budget available for education and public sentiment about the "quality of education" being purchased was critical. Many of these same issues were being experienced by educators in the United States at the same time.

In late 1971, an article by Olson, published in Phi Delta Kappan received considerable attention from the administrative staff and Board members of the Board of Education for the City of Toronto. The Board requested their Research Department to prepare a review of the literature on class size. In response, Shapson (1972) published a brief review of the literature which drew attention to some of the inadequacies of the previous research. At the request of the Board, this publication was advertised for sale in the local press, which had by this time written several articles on the issues surrounding the "class size question," the pupil-teacher ratio, and contract negotiations with teachers. About a month later, the Toronto Teachers' Federation approached the Board with a request to study the effects of class size, since existing literature was ambiguous. The Board then requested the Research Director to prepare a proposal and draft a budget

for the study of the effects of class size. As this proposal was found by the Board to be beyond its means, it was forwarded, at their request, to the Minister of Education for consideration as contract research. A contract was prepared and signed in October 1973.

The preceding has provided a brief overview of the roots of the study and its raison d'être; the rest of this section will consider the design and implementation problems encountered by the research team, followed by details of the research management process. It was by means of this process that a study of this scale could be completed with minimal disruption to the students, teachers, and administrators, while maximizing the potential of the research design to accomplish its task.

By the time that the study was officially undertaken, a number of problems had been identified both in reviewing the literature and in discussion with administrative staff within the Board. A selection of these problems is presented below:

1. Atypical students should be distributed equally to each class size studied. A method for assigning students to class sizes must include safeguards to prevent accidental or intended biasing of the results of the study by overrepresentation of any student characteristic such as academic ability or proficiency in English in any particular class size. Some studies in the past which have examined existing classes have found better achievement in the larger classes than in the smaller classes (Little, Mabey, & Russell, 1971). Findings such as this may well reflect a tendency on the part of principals to put the better-behaved students in the larger classes and the poorer-behaved students in the smaller classes when a choice is possible.
2. Teacher ability and teaching style ought to be controlled. As no known assessment of teacher ability or teaching style has been accepted by the

profession this must be accomplished by the research design. In conversations prior to designing the study, it seemed that the issue of controlling for teaching ability was the most commonly mentioned.

3. The definition of "small" and "large" class sizes ought to be rational and have practical implications for decision makers.
4. Teaching experience and other background factors such as academic preparation ought to be controlled since they may be a factor in teacher effectiveness. As with (2) above this must be accomplished by the research design.
5. Socio-economic factors relating to the community served by the school system and to the pupils should be observed and reported. This would ensure that no bias to the findings of a study was directly related to economic factors rather than to change in class size alone. For example, within the Toronto public school system, class size is directly related to population characteristics by a staffing formula. In major studies which have included many different communities, with greatly varying expenditures on education, one suspects that the wealthy communities attempt to reduce class size, hire good teachers, provide good educational resources such as extensive libraries, etc. Thus any findings relating to class size could well be confounded by the other variables. As well, the socio-economic characteristics of the sample may limit the generalizability of the results.
6. An attempt should be made to explore as many relationships with the class size variable as are held important by as diverse a population as possible, including students, teachers, parents, administrative staff, trustees, etc. Given the many objectives which these different groups have for education, one must consider many variables if a study is to have any likelihood of receiving serious consideration.

7. The length of exposure to the experimental condition and an adequate description of the circumstances in which the study was conducted ought to be documented. Many previous studies have failed to report that the limits of generalization were to a specific subject matter or to specific time conditions such as the creation of a small class size to teach one subject for a few weeks.
8. An adequate number of teachers and classes ought to be included in the sample to permit generalization to the desired population and to permit the power required for analysis.
9. Student's attitudes and opinions might well be included in a study of class size since they are users/consumers/beneficiaries.
10. Teachers may have to be trained to use techniques appropriate to different class sizes.

PART II

An important characteristic of the contract for this research and the main subject of this paper was that the first year of the study was allocated to design, planning and implementation of the study. As the Principal Investigator was obliged to continue in his role as Director of Research, approximately 15 to 20 per cent of his time was available to the project. The Project Director, S. Shapson, was able to devote 75 to 80 per cent of his time. Two full-time Research Assistants were hired - one responsible for paper and pencil instrumentation and one responsible for observation protocols. The Principal Investigator and Project Director negotiated the implementation of the study with various officials and administrators within the Board and with their counterparts in both the North York and Scarborough Boards of Education.

When the Board originally approved a study, in response to the Federation's request, they asked that the Federation be consulted when the proposed study was initiated. They were approached and a resource group was established. This group helped to refine procedures, to define the scope of the study, to improve instruments and to identify additional areas in which attempts should be made to measure effects. Two major concerns of the resource group seemed to be that the process variables would not be thoroughly evaluated and that outcomes which related to "creativity" in composition, art, self-expression, and so on, would be omitted.

Even before the Federation was contacted, two key meetings had been arranged, the first with the Directors of Education in the City and in the adjacent boroughs, the second with the Area Superintendents and the Principal Officials for the Toronto Board of Education.

The meeting with the Directors provided the vehicle for approaching the senior officials in their boards and gave the study an official approval for the Metropolitan area. Of the five borough boards, three agreed to let schools take part. In one of these three boroughs none of the schools which seriously explored ways to participate was able to take part in the study.

In Toronto the administration agreed to ask for at least one participating school from every one of the six administrative areas and was successful.

Another problem was obtaining the additional teachers for the study. Ministry contracts do not provide for additional instructional staff and the extra staff to cover the smaller sizes had to be found within the staffing formulae. This required and received further co-operation from officials within the boroughs and the City.

The most significant change in the initial study plan came at the first meeting with the Area Superintendents. In trying to ascertain what would be the most suitable sizes for the study, the Research Department had made five proposals. These alternatives were discussed and as various problems were considered, one Area Superintendent suggested the sizes which were eventually used. These sizes which were suggested -- 16, 23, 30 and 37, each plus or minus 2 -- were shown in further discussion to be administratively useful, consistent with the plans for the study, and superior to the other suggestions.

Although the issues related to the selection of class sizes were not documented at that stage, some of the features that were considered important at that time are noted here. It was considered essential that there be more than two class sizes so that it would be possible to identify with more confidence any linear trends in the data. In order to identify critical points, the more class sizes and the smaller the steps between them, the easier it would seem to identify critical points. However, in an experimental study one could only plan for a few different sizes. Since the study was being operated in a field setting, and since there would be transfers of students in and out of the schools, it was desirable to keep the size of the step between two class sizes as large as possible so that there would be some variability possible within each size, and so that the adjacent sizes would not overlap. Finally, it was felt that the class sizes should be acceptable to teachers and parents.

The reasons for the attractiveness of this collection of four sizes are relatively easy to specify. The class size of 37 ± 2, provides for sizes from 35 to 39. Size 40 is avoided: it appears to be a size which is considered very undesirable by many parents and teachers. On the other hand, class sizes of 35 are not uncommon and occasionally, where special

circumstances obtain, one will find a school with a class of 37. Nonetheless, a class that has more than 35 students would be considered "too large" by many administrators. This category represents a size that is larger than is commonly found today and is essentially a social and political maximum.

Class size of 30 $\pm$ 2 appears to represent some kind of "reasonable" ideal from an administrative point of view. In recent years, many classes, especially in the lower grades and inner city, have been maintained at this level, or slightly lower. It would seem that this category is considered common and even "normal" and might well be a target as a maximum at which many administrators would aim.

A class size of 23 $\pm$ 2 gives a class size of 25 or smaller, which appears to be seen by some teachers as an ideal. Olson's work on the "Quality of Education," (1972) suggests that 25 represents a critical point; there was a notable increase in the "quality" scores in classes that were smaller than 25.

The class size of 16 $\pm$ 2 may be seen as somewhat impractical, yet closer examination indicates the importance of such a size in the study. This is a size which is maintained for many special education classes. This is the kind of class size which is found in some private schools. Also, from a research point of view, below 20 may be a size where some significant differences in interpersonal relationships and classroom process may be easily and clearly demonstrated.

The T.T.F. resource group had its first meeting without any members of the research team and had available all the materials discussed by the principal officials. It was reported by the chairman that they also found the sizes of 37, 30 and 23 quite suitable for a study. They initially

questionned whether there was any point to having the size of 16 and apparently after some discussion and examination, decided that the study would be incomplete if classes of that size were not included.

So that the study would not be limited to large schools which could supply 106 students for all four sizes, it was decided to include schools that could provide 53 students; one year they would have sizes 16 and 37, the other year 23 and 30.

Different Area Superintendents arranged different procedures for requesting the participation of a school. Eventually, in each area one or more principals were found who would be willing to consider the project for their school. The Principal Investigator and the Project Director then met with the principal to explain the project, and to discuss any particular problems which the study might create because of the school's characteristics. At all but one school this was followed up by a meeting with the staff. It depended greatly on the principal how this meeting was arranged, and who attended it. Some principals arranged it for all the staff, some for teachers of the junior division. Some principals, because of their general procedures, left the meeting voluntary, whereas others made the meeting compulsory. In one school it was held on a Professional Development Day as part of their program. In most schools it was held after school, but in one or two it was held at noon.

The initial climate of these meetings varied greatly, although there was never any hint of hostility. One interesting meeting occurred in a school where a principal, in an attempt to be co-operative, had provided the teachers with no information about why the staff meeting was called, lest he mislead them, so that the teachers did not know the purpose of the meeting. It was very rare for a principal to have done more than

outline in a few sentences the purpose of the meeting, so that once again at the meetings the project was outlined, details were discussed, questions answered, and particular problems which might apply to a given school were raised by the teachers.

Although it was difficult to ascertain at the conclusion of such a meeting whether or not the teachers would be interested in participating in a study, both the Principal Investigator and the Project Director were pleased by the thoughful approach the teachers took to the project, the interest they displayed in it, the concern for design which was manifested by a few teachers in each school, and the care with which they considered the various aspects of the project and possible implications for them, the students, the parents, and the school. This extended discussion with principal and staff helped anticipate and avoid possible difficulties in individual schools.

At none of these meetings did the Principal Investigator attempt to have teachers volunteer at once, and indeed, for the teachers it might be said that the project was undersold to make sure that all the possible drawbacks which might concern the teachers were identified. This way all the participants had a fair idea of what might be involved in participating in the study and could consider it carefully.

Several of the schools where the study was discussed with the staff were unable to participate in the study. In one, after a warm and enthusiastic reception by the teachers, they decided that parental objections would be insurmountable (this was a school in which parents felt their children had been used as "guinea pigs" in previous years). In other schools, there were difficulties in getting a sufficient number of teachers to participate. The two-year commitment was a major problem for teachers who

hoped to change schools or grades by that time. In one school, participation was not possible because of an unusual distribution of students among the various grades and all attempts at arranging classes failed, in spite of a meeting with the principal, a meeting with the staff, and a further meeting with the principal where all efforts were put towards solving these problems. Because of falling enrolment one principal felt that he would not have enough students to provide all four sizes for two years. His Area Superintendent gave him the responsibility of arranging another school for us to visit to try to make up a pair of schools. As noted before, a school could have the sizes of 16 and 37 in the first year and the sizes of 23 and 30 in the second year, or vice versa. This would ensure that the school had all four sizes within the course of the study, that teachers were rotated through the various sizes, and that the same number of students would be involved for both years of the project. In this and another area an attempt was made to "pair" schools to provide the four sizes for each year. A school with a similar student population (i.e. a similar proportion of non-English speaking homes and a similar socio-economic level) was sought to introduce further balance. There was only one participating school where neither the Principal Investigator nor the Project Director met with the staff before the school's participation was obtained.

As the preceding comments may suggest, the Principal Investigator felt that the schools differed greatly in their general climate. It was his feeling that the schools represented various kinds of "learning environments." The comments of the observers and research technicians who worked in the various schools verify these perceptions. Questionnaire responses from the teachers suggest that only one teacher genuinely did not volunteer for the study. There were two teachers who were not certain whether they had really volunteered or not.

Most of the schools which could not participate in the study continued to show interest in the project and provided a great deal of help by making classrooms available for pilot testing of the observational schedules and other instruments.

The above series of visits and contacts with officials, individual principals and their staffs took much longer to carry out than might appear from the first reading. The initial meeting had been held with senior staff in Toronto early in November. It was not until the beginning of March that the researchers could meet with an individual principal and his or her staff. Not only were the numerous meetings, including in one school a meeting with the parents, time-consuming, but the procedure of making sure that schools were participating on the basis of "informed consent" also made it an anxious four months. It was not until March that there was any certainty that even one school was in the project.

The procedures which have been briefly described above are a significant feature of the study. Very few research studies have had such extensive preliminary discussion at all levels from the Director to senior officials to principals to staff, within a school system before their implementation. Although participation in the project was not a matter of choice for parents and pupils, a letter, with Chinese, Italian and Portuguese translations where necessary, was sent home to every parent explaining the study briefly and promising that no student would be in the large class for two years.

Thus the study began with a total of 12 schools involving 35 classes participating in the study (8 schools from Toronto, 2 from North York, and 2 from Scarborough). The sample size is smaller than was originally planned, but further efforts were unable to add any further schools to the project.

In general, the following several factors played a part in ensuring adequate participation. The principals had to feel that they would get some staffing relief in return for setting up the various sizes. This staffing relief had to be provided in a way that was seen as fair and equitable by other principals. It would appear that in nearly all cases principals who agreed to participate had either neutral or positive attitudes towards research. The support of the Area Superintendents and senior officials in the school system seemed essential. It is likely that the participation of a few principals was based primarily on support of the project by senior officials.

It should be noted that in the study there was a definite over-representation of schools whose pupils came from low-income homes, compared to the City of Toronto population. This seems to be primarily because schools in the lower-income areas tend to be large and therefore it was easier to arrange the study within the organization of such schools. Many of the schools serving the middle-income neighbourhoods had much smaller student populations, primarily because single-family dwellings produce less densely settled neighbourhoods; these schools could not provide enough students for the study.

Assignment of Students and Teachers

The procedures used for the assignment of the students were set up so that it was possible to cope with any peculiarities which might be encountered. There was a card for each student which could be manually shuffled and sorted. Based on the school's own frame of reference, the students were identified - usually by the principal - as either high, average or low relevant to the academic performance of the other students in the school. Students' cards for each school were then divided into six piles based on this "rating" and sex. From these six piles random assignments were made to

the different class sizes so as to keep the proportion of students from each group the same as the proportion of these students in the school.

Using such a stratified random technique increases the confidence one can have in assigning any differences to the effects of class size. It also imposes more control over the study at the school level and increases its credibility among the teachers at the school. The teachers were aware, of course, that every attempt had been made to make each group "equal." After the differently sized classes were formed for a school, the teachers were randomly assigned to the classes in that school.

The procedures which were followed made it possible to keep two students away from each other or to make a specific placement of a student away from a teacher if it were required. The schools had to specify any special requirements, e.g., "You mustn't let Johnny and Frank be in the same room because they are twins," or "Mary had Mr. Brown in Grade 2 and she shouldn't have him again," before the random assignments were made by the researchers. In these cases, after the stratified random assignment of students had been accomplished, the files were checked to see if any specific requests, or other conditions had been violated. If chance had kept them apart nothing was done -- otherwise, an exchange was made with a student from the adjacent size in the same stratum.

The actual number of specific requests over the two years of the study could be counted on the fingers of both hands (and only once or twice was an alteration required) but the possibility of having such wishes taken into account once again seemed to enhance the study in the eyes of the school personnel.

The sorted cards were used to generate class lists. In the whole study there were two exceptions to the complete procedure. One exception, found in two schools, was to have the random assignment of teachers to the first large or small class accomplished by lot draw in the principal's office. The second exception was in a school where the principal commonly followed sequential assignment of students after they had been arranged in descending levels of ability. The principal was allowed to follow his usual procedure under the Research Department's supervision the first year.

In the second year of the study, a similar procedure was followed, except now there was one more constraint, namely that students who had been in a class size of 16 the first year could not be in one the second year, and students who had been in a class size of 37 the first year had to be in one of the three smaller sizes the second. Once again, random assignments were made using the strata based on the rating of the students' academic standing and sex. Exchanges were made when the above condition was not met.

Teachers were assigned to these classes randomly but here too there was an additional constraint -- a teacher who had taught a size of either 30 or 37 the first year had to teach a size of 16 or 23 the second year, and vice versa. This meant that in most classes in the second year the teacher had a few students whom he or she had had the first year.

At the conclusion of data collection, there were a total of 38 teachers who had been involved in the study for part or all of it. In several cases, there were classrooms which were dropped from the analyses because of problems that related to these staff changes. A total of 30 teachers were in the study for both years. The analyses are based on 62 classes -- 30 in the first year, 32 in the second year. The actual size of the classes was monitored eight times during the course of the year to

be sure that the size constraints were satisfied. In one case, variation from the assigned class was so great and occurred so early in the year, that it had to be reassigned (this was a class size of 37 that began with 35 pupils registered, dropped to 33 pupils by late November, and never had more than 33 after that). Two other classes fell outside the $\pm$ 3 range in the latter half of the year, but it was decided to keep them within their regular class size assignment.

PART III

It is probably obvious by now that the present study was not the idea study of the effects of class size. Compromises had to be made which limited the validity and generalizability of the results. However, attempts were made to deal with some of the problems discussed in the first part of this paper, and the authors tried to take the study's shortcoming into account when drawing their conclusions. In order to describe the limitations and the generalizability of the study, a review of how its design dealt with the problems discussed in the first part of this paper will now be presented.

(1) <u>The over-representation in small classes of atypical students</u>.

The stratified sampling procedure was intended to prevent the small classes from being overloaded with atypical children. Although students were stratified only by academic ability, the random sampling within strata was another control intended to prevent the assignment of children with non-academic problems exclusively to one class size.

A few requests were made by schools to be allowed to assign a child with special problems to the smallest class size. The researchers agreed to only one of these requests; for this request reasons other than

simply providing the child with the supposed benefits of a small class were given. While the authors may have been unaware of some assignments of problem children to a particular class size, it is unlikely that such assignments, if there were any, were frequent enough to confound the experimental design.

Some control of this type is necessary in any study of class size.

(2) Teachers' abilities and teaching styles must be controlled.

It was noted in the first part of this paper that there is little agreement on definitions of teaching style or teachers' abilities. As a result, assessments of teachers' abilities and teaching style cannot be used as controls (in an analysis of covariance, for example).

Four types of control were attempted in the present study. The first control was to randomly assign teachers to class sizes -- a condition that was met without exception. (The importance and feasibility of random assignments in a field setting is discussed in an excellent recent article by Boruch (1976).) A second and additional control was the random assignment of teacher to two different sizes. This was intended to balance the effects of individual teaching styles; however, because there were four class sizes such effects may still remain confounded with class size, or with combinations of class sizes.

The third type of control was to reanalyze the data using a balanced design. By using paired t-tests to compare teachers' scores in large classes (that is, classes of 30 or 37) and small classes, teacher effects could be controlled. However, the idea of the study was to compare four class sizes, and treating the four sizes as if they were only two seems somewhat beside the point. As well, the power of the statistical test is dependent on the correlation between teachers' scores, and will vary from variable to variable.

The fourth type of control was the use of multiple linear regression with dummy variables to remove teacher effects from scores; the residual scores were then analyzed by analysis of variance. The problem here is that if some teachers adapt to class size and other teacher don't, class size effects may be removed along with the teacher effects. And again, power may vary from one variable to another.

Despite the shortcomings of the controls, it is unlikely that any of the class sizes was taught by a group of teachers which differed extremely from the groups of teachers teaching the other class sizes; however, control of irrelevant variance is one of the defining characteristics of experimental research, and the necessity (and the difficulty) of providing such control cannot be glossed over.

(3) Definition of class size.

Ideally the definitions of large and small classes would be based on a consideration of previous findings and relevant theory; the definitions used in the present study were not ideal. First of all, there is no scientific theory of class size effects, so that definitions of large and small classes cannot be derived from one. Secondly, definitions of large and small classes had to be administratively acceptable -- that is, the schools had to be able to provide large and small classes within the constraints imposed by the budgeting formula. However, the class sizes used in the present study did have some relationship to previous findings and to current class sizes. As was noted, the largest classes in Toronto are about the size of the largest classes in the present study. The class size of 30 is close to the average class size in the junior grades in Toronto, while the class size of 23 is just below one of Olson's (1971) so-called breakpoints (a class size below which the "quality of education" is supposed to improve markedly). Finally, the class size of 16 seems to be thought of as a

pedagogically ideal size -- the restriction of special education classes in Toronto to this size implies that classes of this size have especially beneficial effects. The sizes selected for this study had the advantage of having been selected after consideration of their relevance to education. However, scientific definitions of class sizes will probably remain a fond hope.

(4) Teaching experience has not been controlled.

Teaching experience was controlled in the same way as teachers' abilities and teaching style. Confounding of this variable with class size in the present study has not been checked. The similarity of the distribution of the teaching experience of the teachers in the study to that of all teachers in North York, Scarborough, and Toronto is also unknown. Since teachers must volunteer for such a study, the representativeness of the sample of teachers will also remain a problem.

(5) SES has not been controlled.

An examination of the 1971 census data for the neighbourhoods of the schools in the study revealed that ten of the eleven schools should probably be considered to be in the bottom half of any SES distribution, and that seven of the eleven should be considered to be in the bottom quarter. This, of course, restricts the generalizability of the results.

As was noted earlier, the correlation of a neighbourhood's population density with its SES may prevent schools from neighbourhoods of high SES from taking part in an experimental study where large classes have to be provided. It is unlikely that anything can be done by educational researchers about this problem.

SES did not become confounded with class size, or, at least, any confounding was unimportant. An analysis of the class size distributions has shown that while there were small differences in the proportion of each class size in two groups defined by SES, these differences were not statistically significant. It appears that differences between class sizes cannot be explained as the result of differences in social class.

(6) Only a few variables have been investigated.

The present study cannot be accused of investigating only a handful of variables, and the process used to select variables should have ensured that they would be of interest to a diverse group of people. However, the use of univariate statistics to investigate the hundred or so variables in the study was dangerous in that a large number of spurious differences could have been detected. Since few differences were found, this problem is not all that important; as well, the use of four class sizes in most of the analyses permitted the researchers to dismiss observed differences as irrelevant or possibly spurious if the relationship observed was non-monotonic.

(7) Documentation of the circumstances in which the study was conducted.

Care was taken to record and report as much information as possible about the circumstances under which the research was performed. The arrangements for the participation of schools and teachers are described in the original report and its technical supplement (Wright, Shapson, Eason & FitzGerald, 1977a, 1977b). Characteristics, socio-economic and other, of teachers and students were reported, as were descriptions of how arrangements were made to administer paper-and-pencil, and such aspects of the classes, as the subject being taught while they were being observed, the teaching methods used, the types of work done by the pupils, and the way in which the classes were organized for work.

(8) Adequate numbers of teachers and classes.

For researchers, an adequate number of teachers or classes is one which gives a statistical test the desired power. The problem of statistical power in the present study was discussed in the original report. This will continue to be a problem, as teachers must volunteer for research of this kind. If too few teachers volunteer, the power of the statistical test will be low. In the present study it was possible to find statistical methods which provided adequate power, although these methods were not without problems.

(9) Students' attitudes and opinions may be considered.

Students certainly have opinions about what goes on in their schools, and these opinions should be of interest to educators. The present study found that students' opinions and attitudes were almost completely unaffected by class size; this finding would seem to be important in discussing the class size issue. In fact, students' opinions would seem to be important in discussing almost any important issue in education.

(10) Teachers may have to be trained in how to teach different class sizes.

This remains an important problem, although not instructing teachers in how to teach small classes before investigating class size effects is justifiable -- after all, it is unlikely that teachers will receive any special training if class sizes are reduced by government. Furthermore, many people seem to assume that reducing class size will automatically produce improvements in education. In this study teachers were given no direction regarding teaching methods for the different sizes. In this way a "naturalistic" situation was maintained. Bereiter (1976) has prepared a good discussion of how to do a study where smaller classes would be taught in a way which exploits the advantages of small classes. There may also be

better instructional techniques for large classes! If class size research is to continue, it would seem best to investigate the efficacy of different teaching methods in different class sizes.

One lesson to be drawn from the Toronto class size study is that the ideal class size study will never by performed. Certainly some very important controls can be used, but others are impracticable. Still, one cannot assume that research on class size is useless. Conclusions can be drawn from an imperfect experiment if the experiment's limitations are taken into account.

On the other hand, this paper demonstrates that it is possible when studying a complex issue to identify significant factors which might affect the validity of the potential study. By responding to an issue when there is wide-spread interest and concern, it is possible to design and implement a study which is sufficiently well controlled that planning can proceed utilizing the findings. As many people who make decisions are not experienced with the technical issues related to research, it is a responsibility of researchers, particularly when responding to topical issues, to introduce the maximum care possible in the project's design. A secondary outcome manifested in this study and possibly to be found in similar projects is the development of a collection of data gathered under such well described and controlled conditions that additional analyses can be undertaken to better understand the process of schooling.

REFERENCES

Berieter, C. Can class size make a difference? In Doris W. Ryan and T. B. Greenfield (Eds.), Clarifying the Class Size Question. Toronto: Ministry of Education, Ontario, 1976.

Boruch, R. F. On common contentions about randomized field experiments. In G. V. Glass (Ed.), Evaluation Studies, Review Annual, (Vol. 1), Beverly Hills: Sage Publications, 1976.

Carr, W. G. New angle of attack needed on class size research. Nations Schools, 1932, 10, pp. 27-30.

FitzGerald, J. M., Wright, E., Eason, G., Shapson, S. M. Some secondary analysis from a study of the effects of class size: the effects of subject of instruction on the behaviour of teacher and pupils. Paper presented at American Educational Research Association, annual conference, Toronto, March 1978.

Little, A., Mabey, C., & Russell, J. Do small classes help a pupil? New Society, October 21, 1971, pp. 769-771.

Olson, M. N. Ways to achieve quality in school classrooms: some definitive answers. Phi Delta Kappan, 1971, 1(53), pp. 63-65.

Shapson, S. Optimum class size? A review of the literature. Toronto: The Board of Education for the City of Toronto, Research Department, 1972 (No. 114).

Shapson, S., Wright, E., Eason, G., FitzGerald, J. Results of an experimental study of the effects of class size. Paper presented at American Educational Research Association, annual conference, Toronto, March 1978.

Wright, E. N., Shapson, S. M., Eason, G., FitzGerald, J. Effects of class size in the junior grades. Toronto: Ministry of Education, Ontario, 1977, (a).

Wright, E. N., Shapson, S. M., Eason, G., FitzGerald, J. Technical supplement: effects of class size in the junior grades. Toronto: Ministry of Education, Ontario, 1977, (b).